DOI: http://dx.doi.org/10.26483/ijarcs.v16i6.7363

Volume 16, No. 6, November-December 2025

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

REFINED MASK R-CNN MODEL TO SEGMENT MICROSCOPY IMAGES FOR ACCURATE BLOOD CANCER DETECTION

Indhumathi Palanisami
Research Scholar, Department of Computer Science,
Park's College (Autonomous), Tiruppur - 641605,
Tamil Nadu, India

Suresh Nachimuthu
Research Supervisor, Department of Computer Science,
Park's College (Autonomous), Tiruppur - 641605,
Tamil Nadu, India

Abstract: Early detection of Acute Lymphocytic Leukemia (ALL) as well as Multiple Myeloma (MM) is critical for reducing mortality rates. One promising new approach for the early detection of these blood malignancies is the Deep Learning (DL) model. However, in order to provide high-quality microscope images for highly accurate blood cancer detection, certain models do not include data diversity enhancement. In order to generate high-quality microscopic images for the prediction of ALL and MM, DeepBCDnet was developed utilizing Resolution Enhanced and Noise Suppression Generative Adversarial Network (RENS-GAN). However, a segmentation approach is necessary for this model to enhance its accuracy. Its scale-invariant structure ignores spatial variations across receptive fields, leading to misclassification of object edge pixels when using Mask Regional Convolutional Neural Network (Mask R-CNN) for image segmentation tasks. This research proposes R-Mask R-CNN, a Refined Mask R-CNN that fuses deep semantic and shallow high-resolution features in the Region Proposal Network (RPN) as well as Region of Interest (RoI) layers using an attention mechanism and a bottom-up structure. At the pixel level, this model successfully identifies and segments micrographs. By incorporating the bottom-up structure into Mask R-Feature Pyramid Network's (FPN) CNN, the path between the lower and top layers is shortened, leading to better usage of features from the lower layers. To fine-tune pixel-level focus, channel-wise and spatial attention methods apply weights to feature maps. A new semantic segmentation layer takes the place of the earlier fully connected (FC) layer; this layer allows for feature fusion through the construction of an FPN and the summing of backward and forward transmissions of feature maps of identical resolution. This layout enhances the data propagation between layers, which in turn improves the accuracy of detection and segmentation. In order to aid classification during segmentation, the network takes into account receptive fields of varying sizes all at once by combining input from multi-scale feature maps. Mask head structure optimizes feature fusion by adjusting the input image scale. Lastly, the forms of blood cancer (ALL and MM) are classified using Dense Convolutional Neural Networks (DCNNs). Deep Blood Cancer Segmentation and Detection network (DeepBCSDnet) is the entire name of the model. The DeepBCSDnet models outperform the state-of-the-art models in terms of accuracy, with 94.71% and 95.57% correspondingly achieved on the SN-AM Dataset, MiMM_SBILab, and C-NMC datasets, respectively.

Keywords: Blood Cancer, Deep Learning, Mask R-CNN, RENS-GAN, Segmentation

I. INTRODUCTION

White blood cells (WBCs) are crucial for immune defense against infections [1]. ALL and MM are types of WBC cancer. A type of blood cancer that starts in the bone marrow and damages white blood cells is known as ALL [2]. MM is the plasma cell cancer in the bone marrow, disrupting normal blood cell production and immune response [3]. While normal plasma cells fight infections by producing antibodies, cancerous plasma cells multiply uncontrollably, resulting in a decrease in healthy blood cell synthesis and problems caused by aberrant proteins. The symptoms of MM can be alleviated by remissions, but the disease itself is incurable [4]. Early detection and diagnosis is important to reduce the morality rate. Traditional diagnosis involves biopsies as well as imaging but these methods can be invasive, time-consuming as well as may cause discomfort or risk to patients [5].

Microscopic images are used as a non-invasive and efficient alternative for detecting blood cancers like ALL and MM [6]. These images enable detailed analysis of cell morphology to identify abnormal cells, helping doctors detect early cancer signs, assess disease severity, and monitor treatment progress without invasive procedures [7]. However, the subjectivity and variability in human interpretation which can lead to inconsistent diagnoses. Artificial intelligence (AI) specifically Machine Learning (ML) as well as DL can address these issues by automating analysis, reducing human error, and detecting subtle patterns, thereby improving accuracy and efficiency in WBC cancer diagnosis [8]. ML identifies data correlations for predictive modeling but struggles with computational complexity and feature selection from large datasets [9, 10]. In contrast, DL facilitates automated prediction leveraging large-scale data for enhanced diagnosis, prognosis and treatment decisions [11].

Various DL Models have been developed for blood tumor prediction. One such model, an automatic detection model for ALL and MM was developed using a Dense Convolutional Neural Network (DCNN) [12] framework with augmented bone marrow microscopic images. Two augmentation techniques were applied: 90-degree image rotations to enable object recognition in various orientations, and edge detection to extract image boundaries. The DCNN model forecasted ALL and MM after feature selection was done using the Chi-Square test. However, these augmentation methods retained the essence of the original images, limiting data diversity and failing to preserve the structural information crucial for blood cancer detection.

To address these limitations, DeepBCDnet [13] was developed using RENS-GAN which generates high-quality microscopic images for ALL and MM prediction. A discriminator, feature extractor, choosing features, as well as

classification are the several stages that make up the model. Two generators are responsible for noise suppression as well as resolution augmentation, respectively. The generators produce high-resolution RENS images from low-resolution inputs, while the discriminator computes adversarial loss to differentiate original and generated images. VGGNet acts as a feature extractor model to preserve essential features. Also, the Stochastic Variance Reduced Extra Gradient (SVRE) method enhances image quality using variance reduction and extrapolation task. Extracted features are selected using SVM-RFE [14] and fed into the DCNN [11] model for accurate blood cancer prediction, ensuring improved image quality and structural information retention. But, this model needs a segmentation method to better localize cancerous cells and capture morphological variations enabling for more accurate blood cancer predictions.

A CNN that merges target detection as well as semantic segmentation, Mask R-CNN [15] is the foundation of most DL models used to detect WBCs in microscopic images. Faster R-CNN is its ancestor. But Mask R-CNN has a hard time making precise predictions about instance details. Recognizing variations in spatial information across receptive fields of different sizes, its scale-invariant fully convolutional structure disregards these discrepancies. Small receptive fields are more concerned with semantics, whereas large one's value detailed information. This imbalance causes the network to misclassify the pixel values at the cell boundaries, failing to capture the relationships between them effectively.

Hence, in this paper, DeepBCSDnet is developed for efficient microscopic image segmentation for the accurate prediction of different blood cancer types, including ALL, MM, and AML. This model constitutes R-Mask-CNN model which overcomes the limitations of Mask R-CNN-based segmentation and effectively analyzes pixel relationships at cell boundaries focusing more on detailed information to increase the performances rate. Within the RPN and RoI layers, R-Mask R-CNN fuses deep semantic as well as shallow high-resolution features using a bottom-up structure along with attention processes. This model is able to recognize and segment pixel-level microscope images with remarkable accuracy. The bottom-up structure is integrated into the FPN of Mask R-CNN, shortening the path between the lower and upper layers. This allows the top layer to make better use of the characteristics from the lower layers. Both the spatial attention mechanism and the channel-wise attention mechanism are responsible for assigning weights to the various components of the feature maps. In addition, a new semantic segmentation layer takes the role of the FC layer. This layer makes feature fusion easier by building an FPN and adding the forward as well as backward transmissions of feature maps at the identical resolution. The bottom-up structure, when combined with attention mechanisms, greatly increases the propagation of information between layers, greatly improving the accuracy of detection and segmentation. The network's ability to incorporate multi-scale feature map data allows it to handle receptive fields of different sizes, giving segmentation accurate classifications with enough contextual information. To achieve optimal feature fusion, the mask head structure adjusts the input image scale accordingly. Finally, a DCNN [12] is employed to classify blood cancer types like ALL and MM. This model effectively effectively analyzes pixel relationships at cell boundaries focusing more on detailed information to increase the models performances rate.

The following format is used to continue the article: Earlier research is contained in Section II. The suggested DeepBCSDnet algorithm is described in Section III, and its efficiency is evaluated in Section IV. The study is concluded in Section V.

II. LITERATURE SURVEY

Cho et al. [16] developed an image transformer for the prediction and classification of ALL. The collected dataset was pre-processed and augmented to reduce the noises and enhances the data dimensions. Vision Transformer (ViT) was employed for the prediction and classification of ALL. However, this model fails to capture fine-grained patterns in microscopic images reduces its accuracy and precision.

For the purpose of segmenting and classifying white blood cell cancer, Ikechukwu and Murali [17] built a deep CNN model named i-Net. This approach used CNN for image training following feature selection, and UNet with the InceptionV2 model (UNet-INV2) for segmentation. To classify normal and malignant WBCs, i-Net was eventually created with more convolutional layers along with optimized hyperparameters. However, accuracy and recall were lower due to its inability to improve data variety by preserving essential data from microscope images for WBC identification.

In order to identify ALL as well as MM from microscopic images, Baig et al. [18] developed CNN-1 along with CNN-2, which are Hybridized CNN (HCNN) blocks. The dataset was pre-processed and segmented to remove background noise and blurriness and enhances the edges. The images were converted into RGB color-space and greyscale using image intensity adjustment and adaptive histogram equalization. Next, the images were inputted into two concurrent CNNs to extract deep features, and then into SVMs to forecast ALL and MM. But, hyper-parameters were not optimized appropriately declining the accuracy outcomes.

A modified DL model for ALL prediction was created by Sampathila et al. [19]. In this method, the collected images were augmented to increase the data size and to avoid the over-fitting issues. Then, the data were fed into CNN to extract the features and softmax layer of CNN was employed to classify the images a healthy or ALL. But, accuracy was decreased owing to its failure to recognise under-stained cells with little colour contrast between the cytoplasm and background.

Ahmed et al. [20] devised an CNN integrated with RF and XGBoost (CNN-RF-XGB) model for the diagnosis of ALL and its types. This approach isolates WBCs from cellular images using three CNN models that have already been trained. Due to the great similarity between normal and infected WBC, a hybrid technique of CNN-RF as well as CNN-XGBoost was used to extract the deep feature maps. These maps were then classified using RF classifiers and XGBoost. But, this model doesn't extract relevant and exact data from microscope images which lowers the accuracy and precision rate.

The CLR-CXG model was developed by Devi et al. [21] to segment and classify WBCs from microscopic pictures of bone marrow. To segment the images as well as extract the key features, this model uses CLR-CXG. Each feature was implemented separately. Following that, in order to achieve high accuracy and remove bias, Convolutional Leaky RELU with CatBoost (CLRC) was employed. On the other hand,

Convolutional Leaky RELU with XGBoost (CLRXG) was developed for AL or MM classification.

Ramasamy et al. [22] constructed an Optimized DCNN (ODCNN) model for bone marrow cancer detection using image processing. The collected dataset was pre-processed to eliminate the noises. Finally, CAT optimization algorithm was used to optimize the DCNN model for the prediction of bone marrow cancer detection. However, generated mask was not able to cover the target edge which lowers the accuracy and recall performance.

In order to detect leukemia on heterogeneous datasets, Saikia et al. [23] created a BSNEU-net, which stands for Block Feature Map Distortion as well as Switchable Normalization-Based Enhanced Union-net. The approach is made up of four Union Blocks (UB), and each of those blocks has Block Feature Map Distortion (BFMD) with Switchable Normalization (SN). In order to get discriminative features, the UB employs union convolution; BFMD aids in pattern generalization and overfitting reduction. SN layers are added to enhance convergence and generalization by addressing the sensitivity of batch normalization to mini-batch dimension changes. However, SN may increase computational complexity, affecting efficiency in resource-constrained environments or with large datasets.

For leukemia identification, Raghaw et al. [24] proposed CoTCoNet, an improved coupled transformer-convolutional network. To create high-resolution leukocyte samples, it uses a Generative Adversarial Network-driven Cell Synthesizer. To integrate and enhance features, it employs a Graph-based Feature Reconstruction (GraFR) module. To optimize and select features, a population-based meta-heuristic method is used. To classify leukemia or non-leukemia, the Hierarchical Deep Learning Classifier (HDLC) is employed. But, it results in potential computational complexity due to the integration of multiple advanced models components limits the models scalability for large datasets.

III. PROPOSED METHODOLOGY

A brief illustration of the proposed DeepBCSDnet concept is provided in this section. The proposed model's overall layout is shown in Fig. 1.

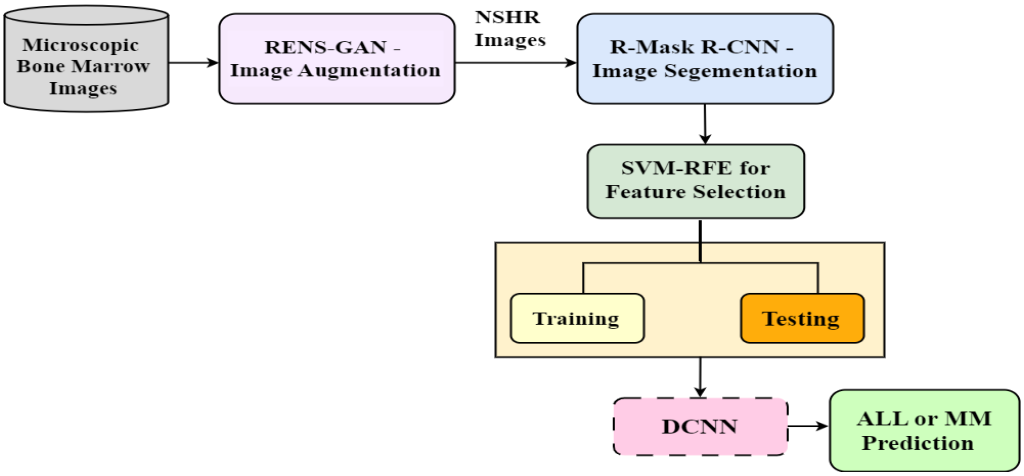

Figure 1. Pipeline of the Proposed Model

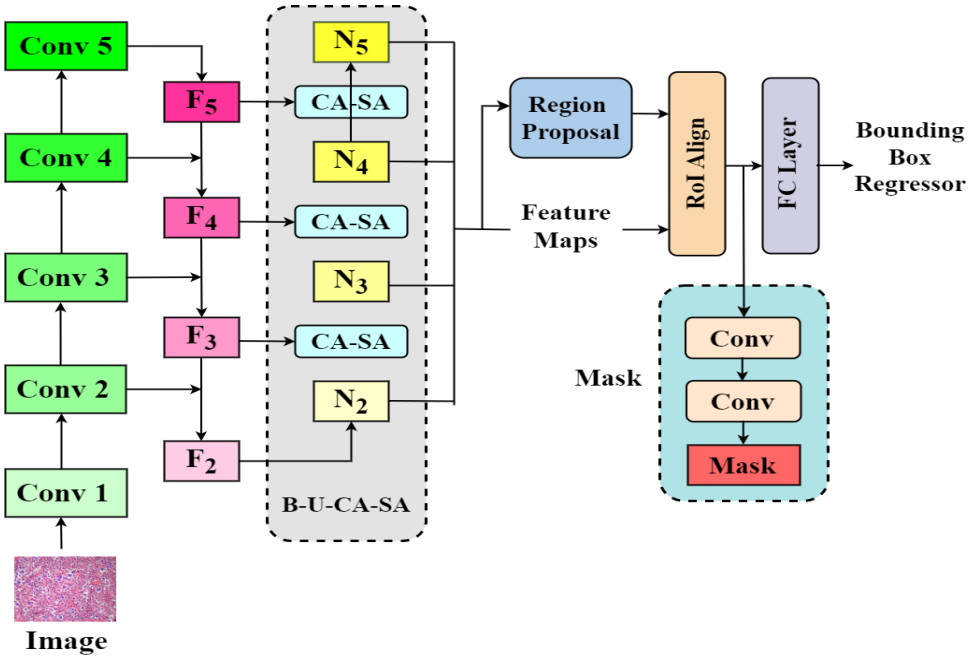

Figure 2. Block Structure of R Mask R-CNN

A. Image Augmentation

The collected images are augmented using RENS-GAN [13] which generates high-quality microscopic images for ALL and MM prediction. There are a number of steps to the model, such as a discriminator, a feature extractor, choosing features, classification, and two generators (one for noise reduction and one for resolution enhancement). The generators produce high-resolution RENS images from low-resolution inputs, while the discriminator computes adversarial loss to differentiate original and generated images. The feature extractor calculates feature loss, which, combined with adversarial loss, forms a joint loss to preserve essential features. In addition, RENS-GAN incorporates the Stochastic Variance Reduced Extra Gradient (SVRE) technique to improve the image quality by reducing variance and enhancing extrapolation via discriminator, generator, and feature extractor mutual training.

B. Proposed R-Mask R-CNN Model for Image Segmentation

The proposed R-Mask R-CNN model captures relationships between pixels at the cell boundaries and emphasizes detailed information for accurate ALL and MM prediction. R-Mask R-CNN uses FPN to improve the Mask R-CNN FPN and create feature pyramids. The bottom-up approach is designed to reduce the amount of data traveling between the lowest and highest levels, allowing for more efficient propagation of data from the lower layers to the top layer. In the bottom-up approach, the feature maps are enhanced to respond to the characteristics of the target by means of channel-wise and spatial attention processes. Fig. 2 shows the block diagram of the R Mask R-CNN model, which uses a bottom-up approach to segmentation using channel-wise as well as spatial attention mechanisms (B-U-CA-SA).

1) Bottom-Up Path Structure

Inspired by the PANet model, bottom-up path structure is developed in this framework. This model incorporates ResNet as well as FPN structures to derive feature maps for four feature levels, specifically $\{F_2, F_3, F_4, F_5\}$. At each feature level N_x $(x = 2, 3, 4, 5)$, the size of the feature map is the same as at the corresponding level F_x $(x = 2, 3, 4, 5)$, and the bottom-up augmented path travels from level F_2 to level F_5.

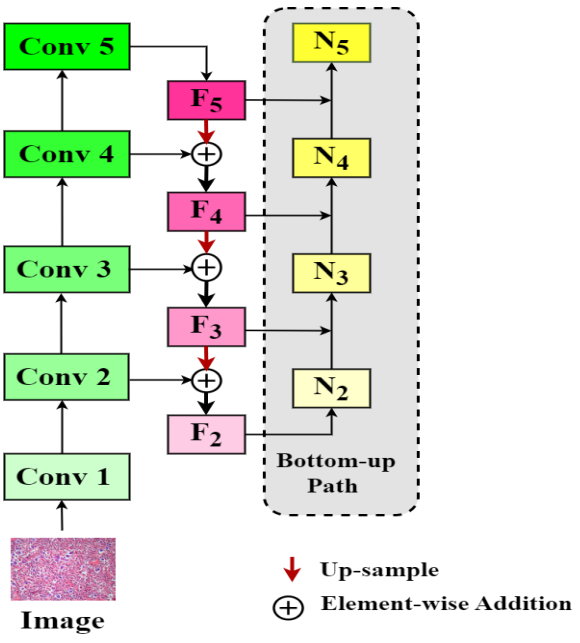

Figure 3. FPN with bottom-up path structure

Then, instead of using $\{F_1, F_2, F_3, F_4\}$, this information is sent to the following network layers through the new feature maps $\{N_2, N_3, N_4, N_5\}$. In Fig.3, the FPN with a bottom-up route structure is shown from N_2 to N_5, with the left side of the figure showing the original FPN structure.

2) Channel-Wise Attention Mechanism

In CNNs, different channels capture varying semantic information, with some channels being more important than others, and some may not contain relevant features. Fig. 4 shows the architecture of the channel-wise attention mechanism.

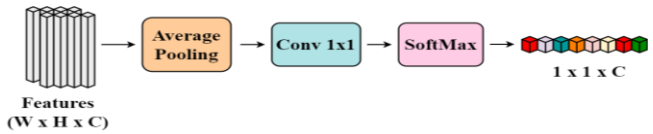

Figure 4. Structure of CA Mechanism

Fig. 4, which displays the F_3 feature maps acquired using convolution, demonstrates that distinct channels possess distinct types of information. Channel 167 and 85 are better for ship-related information, whereas channel 97 is more suited for background information. Most current methods do not use channel selection when extracting features using convolution and instead give equal weight to each channel. By giving more weight to the channels that respond to the target more strongly, the channel-wise attention method improves feature extraction for input objects.

In order to accomplish channel-wise attention, the set of convolution features $C = \{C_1, C_2, \ldots, C_Z\}$ is taken into account, where $C_x = R^{W \times h}$ is the x^{th} channel in the feature map F and Z is the sum of all channels. As shown in Eq. (1), a channel feature vector Q is generated by applying the average pooling function to each feature map C_x.

$$Q = \{Q_1, Q_2, \ldots, Q_Z\}, \quad Q \in R^Z \tag{1}$$

In this context, Q_x represents the features extracted from the x^{th} channel following average mixing. The aggregated features from every channel in Q_x are trained using convolution with a 1×1 kernel. The feature vector Q, which includes all of the channels 1, goes through to the softmax function. A few example features in various channels are shown in Fig. 5.

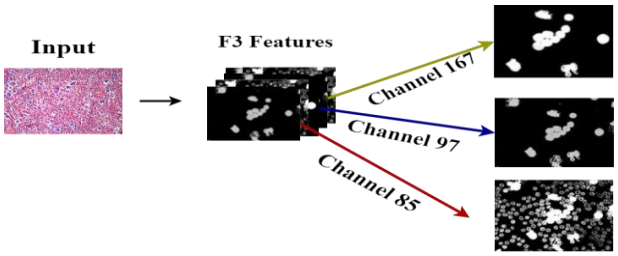

Figure 5. Sample Feature Variation in Different Channels

3) Spatial -Wise Attention Mechanism

In recognizing objects from images, the target objects are usually localized to certain areas of the image instead than being present throughout. The feature map pixels are given weights by the spatial attention mechanism, as opposed to all regions being processed equally by normal CNNs. There will be less background noise because the algorithm is concentrating on the foreground pixels. The spatial attention

mechanism eliminates background interference by assigning higher weights to pixels in the feature map corresponding to foreground objects. This sharpens focus on relevant regions, suppresses irrelevant background details, and enhances edge clarity in images. Take into consideration the convolutional features $C = \{C_1, C_2, \ldots, C_z\}$ and a 1×1 convolution kernel in order to produce the feature map representation G. The feature G's pixel points are subsequently subjected to the softmax operation, with the condition that their values add up to one. Fig. 6 depicts the structure of SA.

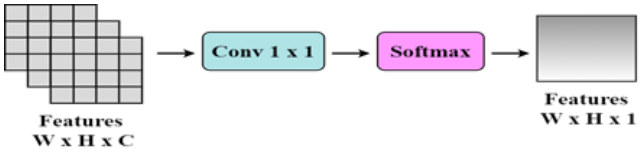

Figure 6. Block Structure of SA

C. *FPN and Bottom-Up Structure with Spatial and Channel-Wise Attention Mechanism*

Below are examples of several configurations that have been implemented in this framework, combining the channel attention (CA) as well as spatial attention (SA) mechanisms (Fig. 8) with the FPN as well as bottom-up structures (Fig. 7).

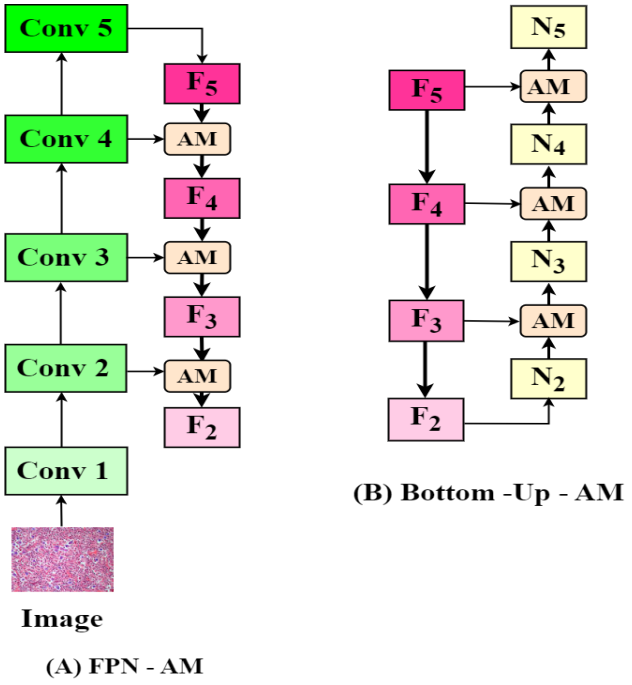

Figure 7. (A) FPN with attention module, (B) Bottom-up structure with attention module.

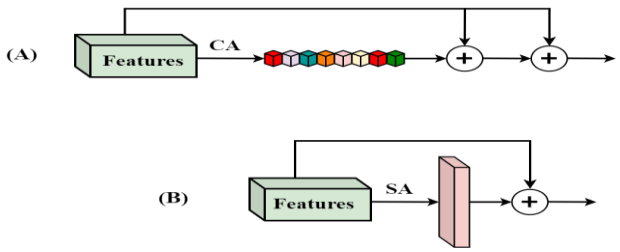

Figure 8. (A) CA Mechanism (B) SA Mechanism

1) FPN with Attention Module (FPN-AM)

Inspired by the PANet model, bottom-up path structure is developed in this framework. The four feature map levels that are first produced using ResNet are $\{P_2, P_3, P_4, P_5\}$. Following these procedures, the FPN is connected to the attention module in order to create the feature pyramid $\{F_2, F_3, F_4, F_5\}$.

Step 1: Create C_5 from feature map P_5 using a 1×1 convolution kernel with 256 channels.

Step 2: In C_2, C_3 and C_4, upsample the feature map C_{x+1} where $(x = 4,3,2)$ by the upsample factor of 2.

Step 3: Merge the upsample feature map with the corresponding P_x, after reducing the channels to 256 using a 1×1 convolution layer

Step 4: In order to create a new feature map C_{pf_x}, add elements one by one.

Step 5: The feature map F_x is obtained by passing C_{pf_x} via the attention module (CA, SA, CA-SA, and SA-CA).

Step 6: Combine F_2, F_3, F_4 and F_5 to form the feature pyramid from the FPN-AM structure.

2) Bottom-Up with Attention Module (B-U-AM)

The architecture of B-U-AM is similar to that of the structure of FPN-Att. In B-U AM, N_2 is same as F_2 (in FPN-AM). The following steps provides the integration of Bottom-up structure with Attention mechanism.

Step 1: Generate feature maps N_{x+1} where $x = 2, 3$ and 4

Step 2: Combine N_{x+1} $(2W \times 2H \times C)$ with a 3×3 convolutional kernel and a stride of 2 using a convolution approach.

Step 3: Minimize the feature maps' dimensions by halving their size with relation to F_{x+1} $(W \times H \times C)$.

Step 4: Integrate the reduced feature maps F_{x+1} through the element wise addition to produce N_x_half

Step 5: In order to get N_x, run N_x_half through the attention module. This can be CA, SA, CA-SA, etc.

Step 6: Combine N_2, N_3, N_4 and N_5 to construct the new feature pyramid $\{N_2, N_3, N_4, N_5\}$

Step 7: Direct the feature pyramid to the subsequent network layers for further processing.

Fig.9 and Fig.10 depicts the block structure of FPN-AM as well as B-U-AM.

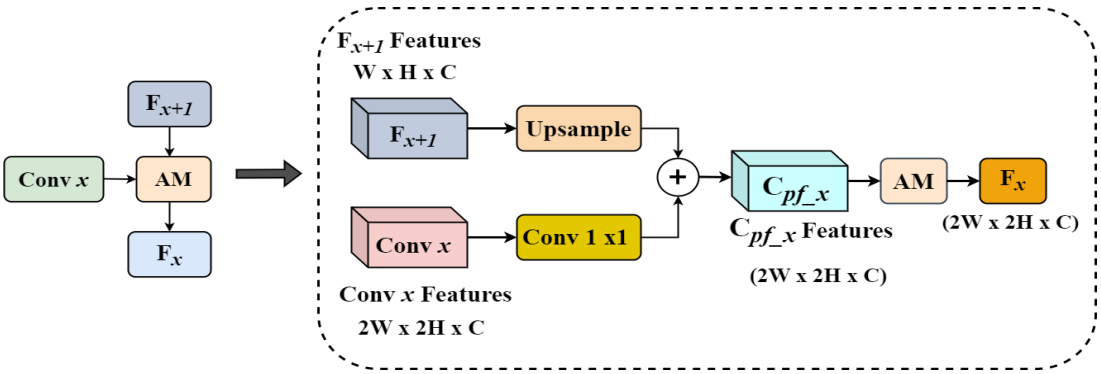

Figure 9. Block Structure of FPN-AM

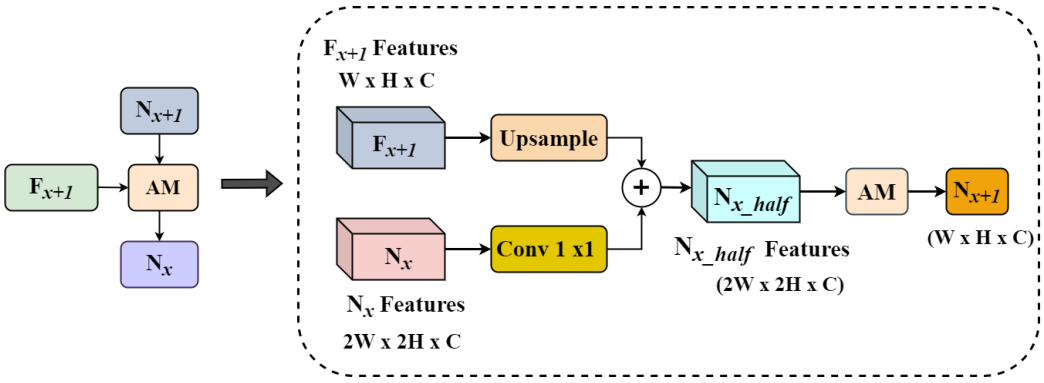

Figure 10.Structure of Bottom-Up-AM

This model effectively analyzes pixel relationships at the cell boundaries, emphasizing fine-grained details to enhance the model's performance. Below algorithm determines the proposed DeepBCSDnet model for blood cancer detection.

Algorithm: DeepBCSDnet model for efficient blood cancer prediction

Input: Microscopic Bone Marrow Images from the Collected Datasets

Output: Segmented feature map and predicted blood cancer type

1. Input the collected images into RENS-GAN model for the data augmentation.
2. Extract feature maps from the input image using the R-Mask-CNN model.
3. Fuse the extracted feature maps using the FPN.
4. Apply a bottom-up structure to the fused features to enhance feature utilization.
5. Perform Channel-wise Attention to assign weights to the feature channels based on relevance
6. Execute Spatial Attention to focus on relevant pixels in the feature maps (foreground).
7. Implement the Semantic Segmentation Layer to refine the segmentation results using feature fusion.
8. Use the DCNN model to classify the segmented features into blood cancer types such as ALL, MM, or AML.
9. Return the segmented features and the predicted cancer type.

IV. RESULT AND DISCUSSION

A. Dataset Description

Three datasets have been used for the experimental purposes.

SN-AM Dataset [25]: Using a Nikon Eclipse-200 microscope at 1000x magnification, images were taken of bone marrow aspirate slides from patients with B-lineage ALL (B-ALL) as well as MM. The slides were stained with Jenner-Giemsa stain. There are a total of 170 images (90 B-ALL and 100 MM) that were shot in raw BMP format and have dimensions of 2560 x 1920 pixels. Any stain normalization methods could be thoroughly tested on MM and B-ALL images due to the high degree of image variability.

MiMM_SBILab Dataset [26]: Images were obtained at 1000x magnification using a Nikon Eclipse-200 microscope and a digital camera. The slides were stained with Jenner-Giemsa stain and used to examine bone marrow aspirate samples from MM patients. The dataset includes 85 images that were shot in raw BMP format and have dimensions of 2560 x 1920 pixels. These 85 images have been stain-normalized using our proprietary technology and are ready for cell segmentation.

C-NMC dataset [27]: A digital camera and a Nikon Eclipse-200 microscope were used to acquire images at a magnification of 1000x. In order to analyze bone marrow aspirate samples from patients with MM, the slides were stained using Jenner-Giemsa stain. The dataset contains 85 2560 x 1920 pixel images captured in raw BMP format. In preparation for cell segmentation, our unique technique has stain-normalized 85 of these images. By combining these three dataset, 7362 images are determined for ALL and 185 images for MM.

B. Performance Evaluation

This part assesses the robustness of DeepBCSDnet model on three different datasets described in Section 4.1. From the collected dataset, two classes ALL and MM are considered. There are 180 MM images and 176 ALL images in the initial dataset. In order to address the problem of data overfitting and improve data variety, 268 photos of ALL as well as 261 images of MM were processed before the experimental task. In the experiment, 70% of the photos were used for training (188 ALL images along with 183 MM images), whereas 30% were used for testing (80 ALL images as well as 78 MM images). The tests are conducted on a Windows 10 (64-bit) laptop that has an Intel® CoreTM i5-4210 CPU (3GHz), 4GB of RAM, and a 1TB hard drive. All models including existing and proposed ones are implemented in MATLAB 2019b to evaluate performance for the two classes.

Parameter Settings: A batch size of 1 was utilized during training, along with an initial learning rate of 0.001 as well as a weight decay of 0.0001. The experimental results show that the optimization features of the TensorFlow framework are well-suited to a lower learning rate, since it allows for faster convergence. With a velocity of 0.7, Stochastic Gradient Descent (SGD) was the optimizer that was used. Every level in the RPN network had one-scale anchor. The five scales $\{16^2, 32^2, 64^2, 128^2, 256^2\}$ corresponded to the levels $\{N^2, N^3, N^4, N^5, N^6\}$, with N^6 being derived from max pooling on N^5. Each anchor had aspect ratios of $\{1:1, 1:2, 2:1, 2:2\}$. The following metrics are used to assess the model's efficacy in detecting different forms of blood cancer:

Intersection over Union (IoU): The region of the union is represented by the area of overlapped or forecasted labels. For an approximation of the predicted boundary's enclosure with the ground truth, it finds the overlap consisting of two boundaries.

$$IoU = \frac{Region\ of\ overlap\ (A \cap B)}{Region\ of\ union\ (A \cup B)} \times 100 \quad (2)$$

In Eq. (2), A is ground truth and B is the predicted segmentation map.

Pixel Accuracy (PA): It is defined as the accuracy of pixel-wise prediction as:

$$PA = \frac{\sum_{i=0}^{K} p_{ii}}{\sum_{i=0}^{K} \sum_{j=0}^{K} p_{ij}} \quad (3)$$

The predicted pixels as class i and the number of pixels from class i predicted as class j are denoted as p_{ii} and p_{ij}, respectively, in Eq. (3), where K denotes the overall number of pixels in the test microscopic images.

Root Mean Square Error (RMSE): Determining the correctness of segregation by evaluating the square root of the MSE value is done as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i} \sum_{j} (S_{ij} - I_{ij})^2} \times 100 \quad (4)$$

The values of i and j represent pixels within the samples, A represents the original sample, S represents the partitioned sample, and N represents the total number of samples in Eq. (4).

Accuracy: It represents the percentage of examined samples that were appropriately labeled and serves as an indicator of the classification system's success rate.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

With a True Positive (TP) in Eq. (5), the classifier has correctly identified an individual with ALL/MM. A correct diagnosis of absence of ALL or MM is indicated by a True Negative (TN).

Both False Positives (FP) and False Negatives (FN) can arise in a classifier. FP occurs when the classifier improperly labels someone as having ALL\MM, while FN happens when it fails to detect someone with ALL/MM and falsely labels them as not having it.

Precision: It is measured by Eq. (6)

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall: This is quantified by Eq.(7)

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1-score: It is calculated by Eq. (8)

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

Area Under Curve (AUC): For each potential detection test cut-off point, the Receiver Operating Characteristic (ROC) curve compares the TP rate (TPR) from Eq. (9) to the FP rate (FPR) from Eq. (10) to calculate an AUC Score, which can take on values between 0 and 1.

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$FPR = \frac{FP}{FP + TN} \quad (10)$$

Fig. 11 shows the results of the suggested model for ALL and MM blood cancers using all of the microscopic images of WBCs.

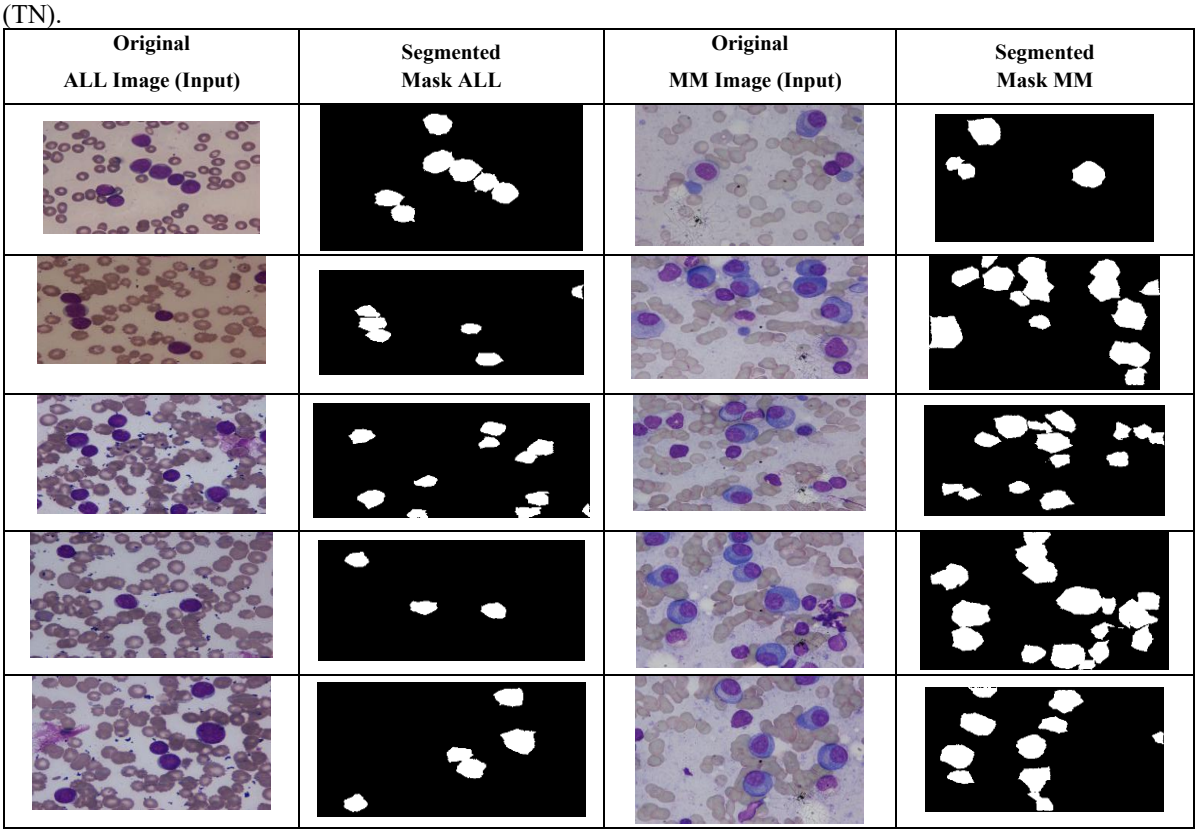

Figure 11. Examples of segmented outcomes using proposed model

Compared to other segmentation models, such as FCM [28], Mask R-CNN [15], and UNet-INV2 [17], the values of IoU, PA as well as RMSE for the suggested R Mask R-CNN are shown in Table 1. This analysis indicates that the R Mask R-CNN achieves a maximum IoU and PA and a minimum RMSE than the other classical models method to capture the relationships between pixels at the cell boundaries in microscopic images for blood cancer prediction.

Table I. Performance analysis for proposed and existing segmentation models

Metrics\Models	IoU (%)	PA (%)	RMSE (%)
FCM	75.29	81.63	37.42
Mask R-CNN	80.56	85.57	31.73
UNet- INV2	84.98	90.34	23.96
R Mask R-CNN(Proposed)	90.03	94.71	15.54

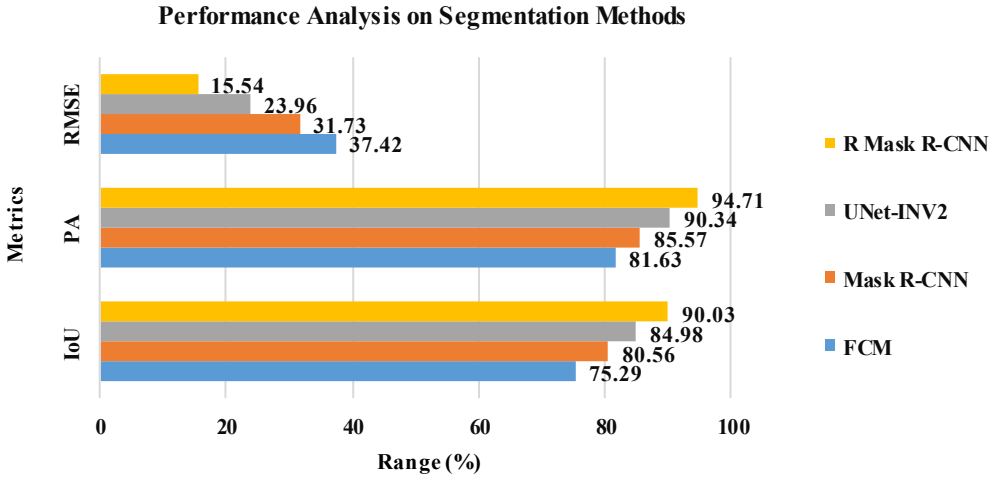

Figure 12.Evaluation of different microscopic image segmentation models

Fig. 12 demonstrates the IoU, PA and RMSE results obtained by the different microscopic images segmentation models tested on the images that collected from three datasets. R-Mask R-CNN improves segmentation by integrating a bottom-up structure with attention mechanisms, enhancing feature fusion for better pixel-level accuracy. Its new semantic segmentation layer and FPN optimize information flow, while multi-scale fusion and an adaptive mask head boost detection and performance. The performance analysis reveals significant improvements with the proposed R-Mask R-CNN. In terms of IoU, R-Mask R-CNN reaches 90.03%, surpassing FCM, Mask R-CNN, and UNet-INV2 by 19.58%, 11.76%, and 5.94%, respectively. For Pixel Accuracy (PA), R-Mask R-CNN achieves 94.71%, outperforming FCM, Mask R-CNN, and UNet-INV2 by 16.02%, 10.68%, and 4.84%. In RMSE, R-Mask R-CNN reduces the error to 15.54%, showing a remarkable decrease of 58.47%, 51.02%, and 35.14%, significantly outperforming the other models.

Table II. Confusion Matrix for the proposed DeepBCSDnet model (Testing images)

Actual/Predicted	Predicted ALL	Predicted MM
Actual ALL	77 (TP)	3 (FN)
Actual MM	4 (FP)	74 (TN)

A study was conducted for two classes (ALL and MM) by combing the three dataset and the confusion matrix as provided in table 2 to evaluate the proposed DeepBCSDnet model with existing models for testing images.

Table 3 exhibits the efficiency of proposed DeepBCSDnet model with and existing classification models including HCNN [18], ODCNN [22], BSNEU-net [23] and DeepBCDnet [13]. Based on the results, DeepBCSDnet outperforms the other traditional algorithms when it comes to tumor class classification.

Table III. Comparison of classification efficiency for proposed and existing models

Metrics	Classes	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
HCNN	ALL	82.74	82.5	81.48	81.99
	MM	80.56	80.77	81.82	81.28
	Average	**81.65**	**81.64**	**81.66**	**81.64**
ODCNN	ALL	86.97	86.25	85.19	85.72
	MM	83.91	84.61	85.71	85.12
	Average	**85.44**	**84.43**	**85.45**	**85.44**
BSNEU-net	ALL	88.67	87.50	88.61	88.05
	MM	87.27	88.46	87.34	87.90
	Average	**87.97**	**87.98**	**87.97**	**87.96**
DeepBCDnet	ALL	93.77	93.75	90.36	92.02
	MM	89.76	89.74	93.33	91.50
	Average	**91.77**	**91.75**	**91.85**	**91.76**
DeepBCSDnet	ALL	96.27	96.25	95.06	95.56
	MM	94.86	94.87	96.10	95.48
	Average	**95.57**	**95.56**	**95.58**	**95.56**

Fig. 13 demonstrates the average evaluation ALL and MM classes for proposed and existing classification models tested on SN-AM Dataset, MiMM_SBILab and C-NMC datasets. It indicates that the DeepBCSDnet model achieves a maximum accuracy in classifying the blood cancer types than the other classical detection models. This is because the multi-scale feature fusion and an adaptive mask head structure enable better context awareness, leading to more accurate classification of blood cancer types like ALL as well as MM. The accuracy of DeepBCSDnet is %, %, % and % greater than the HCNN, ODCNN, BSNEU-net and DeepBCDnet models.

The accuracy of DeepBCSDnet is % higher than HCNN, % higher than ODCNN, % higher than BSNEU-net and % greater than the DeepBCDnet Models. Similarly, the recall of DeepBCSDnet is % higher than HCNN, 2% higher than the ODCNN, 7.88% higher than the BSNEU-net and % higher than the DeepBCDnet models. The F-measure of DeepBCSDnet is %, %, % and % larger than HCNN, ODCNN, BSNEU-net and DeepBCDnet models. Accordingly, it summarizes the DeepBCSDnet model outperforms the conventional models to segregate the blood tumor (ALL and MM) regions and classify the tumor classes appropriately.

Fig.14 illustrates the AUC of different classifiers to identify and categorize blood cancer types like ALL and MM. It indicates that the AUC of DeepBCSDnet is which is 0.94 which is 54.09%, 28.77%, 16.05% and 4.44% higher than the HCNN, ODCNN, BSNEU-net and DeepBCDnet models. The reason behind this is that a semantic segmentation layer, which takes the place of the FC layer, improves the representation of fine-grained features and reduces noise and false positives by directing attention processes to relevant regions.

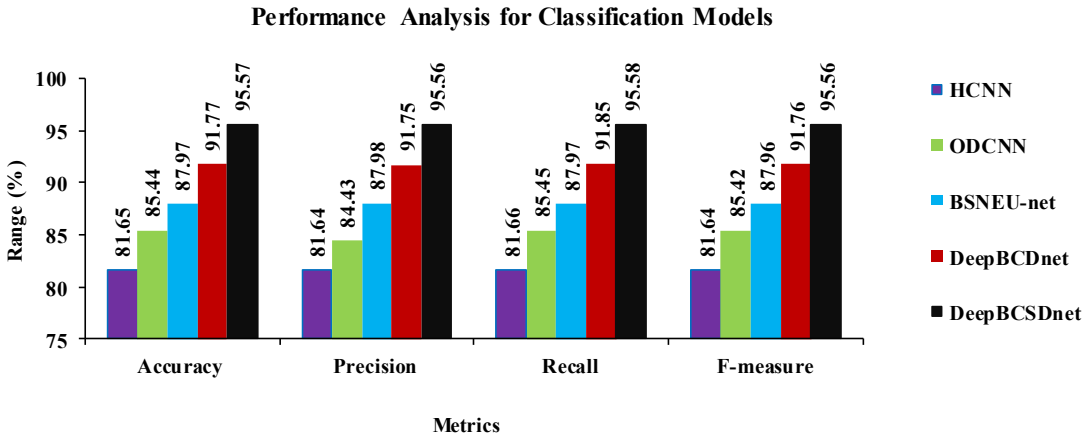

Figure 13.Statistical analysis of different blood cancer classification models

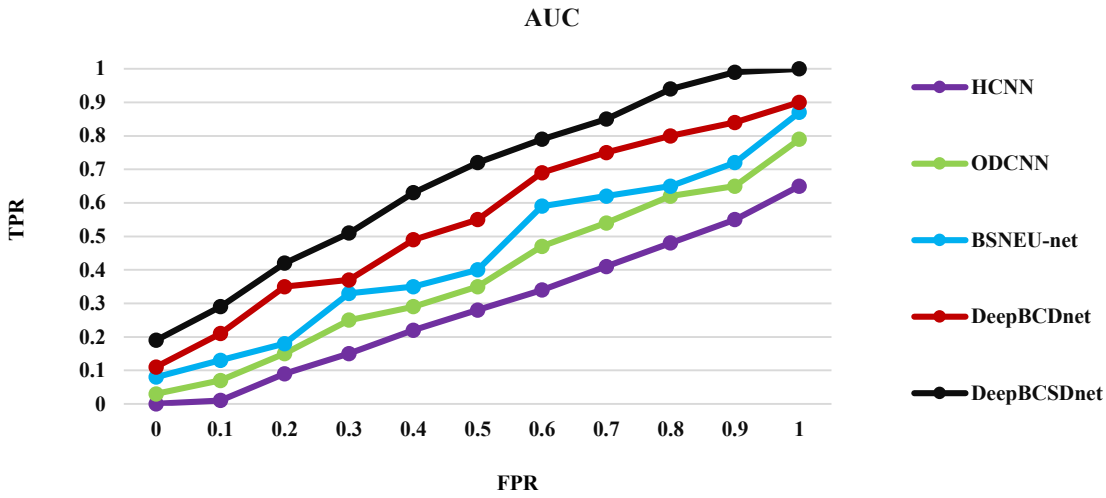

Figure 14.AUC compassion of the Proposed and Existing Models

V. CONCLUSION

In this paper, DeepBCSDnet model is suggested for efficient microscopic image segmentation model for predicting blood cancer types like ALL, MM and AML. It enhances Mask R-CNN by integrating a bottom-up structure as well as attention mechanisms within the RPN and RoI layers, improving feature fusion and pixel-level segmentation accuracy. The bottom-up structure in the FPN shortens pathways between layers, enhancing the use of lower-layer features. CA and SA assign weights to features and pixels, respectively, while a new semantic segmentation layer replaces the FC layer for better feature fusion. Multi-scale feature map fusion provides contextual information for improved classification. Finally, a DCNN classifies blood cancer types, ensuring accurate segmentation and prediction. At last, the test findings proved that the DeepBCSDnet model on the collected databases has 94.71% and 95.57% of accuracy for segmentation and classification, respectively compared to the other conventional models.

VI. REFERENCES

[1] A. Bagramyan, J. W. Wu, K. Mirsanaye, C. Alt, and C. P. Lin, "Towards non-invasive white blood cell count in humans," medRxiv, 2024-11, 2024.

[2] J. H. Yoon and S. Lee, "Diagnostic and therapeutic advances in adults with acute lymphoblastic leukemia in the era of gene analysis and targeted immunotherapy," The Korean Journal of Internal Medicine, vol. 39, no. 1, p. 34, 2024.

[3] D. Fotiou, M. Gavriatopoulou, and E. Terpos, "Multiple myeloma and thrombosis: prophylaxis and risk prediction tools," Cancers, vol. 12, no. 1, p. 191, 2020.

[4] M. S. Abduh, "An overview of multiple myeloma: a monoclonal plasma cell malignancy's diagnosis, management, and treatment modalities," Saudi Journal of Biological Sciences, vol. 31, no. 2, p. 103920, 2024.

[5] A. Almadhor, U. Sattar, A. Al Hejaili, U. G. Mohammad, U. Tariq, and H. Ben Chikha, "An efficient computer vision-based approach for acute lymphoblastic leukemia prediction," Frontiers in Computational Neuroscience, vol. 16, p. 1083649, 2022.

[6] V. Patakvölgyi, L. Kovács, and D. A. Drexler, "Artificial neural networks based cell counting techniques using microscopic images: a review," in Proc. IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI), May 2024, pp. 327–332.

[7] M. Shahzad, F. Ali, S. H. Shirazi, A. Rasheed, A. Ahmad, B. Shah, and D. Kwak, "Blood cell image segmentation and classification: a systematic review," PeerJ Computer Science , vol. 10, p. e1813, 2024.

[8] O. Islam, M. Assaduzzaman, and M. Z. Hasan, "An explainable AI-based blood cell classification using optimized convolutional neural network," Journal of Pathology Informatics, vol. 15, p. 100389, 2024.

[9] C. Ananth, P. Tamilselvi, S. A. Joshy, and T. A. Kumar, "Blood cancer detection with microscopic images using machine learning," in Machine Learning in Information and Communication Technology: Proc. ICICT 2021, Singapore: Springer Nature, pp. 45–54, 2022.

[10] H. Singh and H. Kaur, "A systematic survey on biological cell image segmentation and cell counting techniques in microscopic images using machine learning," Wireless Personal Communications, vol. 137, no. 2, pp. 813–851, 2024.

[11] S. Glüge, S. Balabanov, V. H. Koelzer, and T. Ott, "Evaluation of deep learning training strategies for the classification of bone marrow cell images," Computer Methods and Programs in Biomedicine, vol. 243, p. 107924, 2024.

[12] D. Kumar, N. Jain, A. Khurana, S. Mittal, S. C. Satapathy, R. Senkerik, and J. D. Hemanth, "Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks," IEEE Access, vol. 8, pp. 142521–142531, 2020.

[13] P. Indhumathi and N. Suresh, "Resolution enhanced and noise suppression generative adversarial network for improving white blood cancer prediction," International Journal of Intelligent Engineering and Systems, 2025.

[14] X. Zhou and J. Wang, "Feature selection for image classification based on a new ranking criterion," Journal of Computer and Communications, vol. 3, no. 3, pp. 74–79, 2015.

[15] M. P. Paing, A. Sento, T. H. Bui, and C. Pintavirooj, "Instance segmentation of multiple myeloma cells using deep-wise data augmentation and Mask R-CNN," Entropy, vol. 24, no. 1, p. 134, 2022.

[16] P. Cho, S. Dash, A. Tsaris, and H. J. Yoon, "Image transformers for classifying acute lymphoblastic leukemia," in Med. Imaging 2022: Computer-Aided Diagnosis, vol. 12033, pp. 647–653, Apr. 2022.

[17] A. V. Ikechukwu and S. Murali, "i-Net: a deep CNN model for white blood cancer segmentation and classification," Int. J. Adv. Technol. Eng. Explor., vol. 9, no. 95, pp. 1448–1464, 2022.

[18] R. Baig, A. Rehman, A. Almuhaimeed, A. Alzahrani, and H. T. Rauf, "Detecting malignant leukemia cells using microscopic blood smear images: a deep learning approach," Applied Sciences, vol. 12, no. 13, p. 6317, 2022.

[19] N. Sampathila, K. Chadaga, N. Goswami, R. P. Chadaga, M. Pandya, S. Prabhu, and S. P. Upadya, "Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images," Healthcare, vol. 10, no. 10, p. 1812, Sep. 2022.

[20] I. A. Ahmed, E. M. Senan, H. S. A. Shatnawi, Z. M. Alkhraisha, and M. M. A. Al-Azzam, "Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features," Diagnostics, vol. 13, no. 6, p. 1026, 2023.

[21] T. G. Devi, N. Patil, S. Rai, and C. P. Sarah, "Segmentation and classification of white blood cancer cells from bone marrow microscopic images using duplet-convolutional neural network design," Multimedia Tools and Applications, vol. 82, no. 23, pp. 35277–35299, 2023.

[22] M. D. Ramasamy, R. K. Dhanaraj, S. K. Pani, R. P. Das, A. A. Movassagh, M. Gheisari, and S. Banu, "An improved deep convolutionary neural network for bone marrow cancer detection using image processing," Informatics in Medicine Unlocked, vol. 38, p. 101233, 2023.

[23] R. Saikia, R. Deka, A. Sarma, and S. S. Devi, "BSNEU-net: block feature map distortion and switchable normalization-based enhanced union-net for acute leukemia detection on heterogeneous dataset," Journal of Imaging Informatics in Medicine, pp. 1–28, 2024.

[24] C. S. Raghaw, A. Sharma, S. Bansal, M. Z. U. Rehman, and N. Kumar, "CoTCoNet: an optimized coupled transformer-convolutional network with an adaptive graph reconstruction for leukemia detection,"Computers in Biology and Medicine, vol. 179, p. 108821, 2024.

[25] https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52757009.

[26] https://www.cancerimagingarchive.net/collection/mimm_sbilab/.

[27] https://www.cancerimagingarchive.net/collection/c-nmc-2019/.

[28] A. Das, A. Namtirtha, and A. Dutta, "Fuzzy clustering of acute lymphoblastic leukemia images assisted by eagle strategy and morphological reconstruction," Knowledge-Based Systems, vol. 239, p. 108008, 2022.

